

Stock Market Analysis and Prediction using Data Science

Shani Kumar

Bellevue University - DSC500-T304 Introduction to Data Science (2195-1) - Spring 2019

Introduction

The Stock Market is a complex world in itself, and the stock price movement is hard to predict. Accurate predictions on how stocks are going to perform remain a crucial issue in finance. Participants in trading are constantly looking for techniques that can better predict stock price movements. Technical and Fundamental indicators based on the past performance of stocks can help us understand how the stocks are going to perform. However many factors can impact market sentiments like political news, social media buzz, and the latest developments in the company are a few of the examples which can change the stock trends.

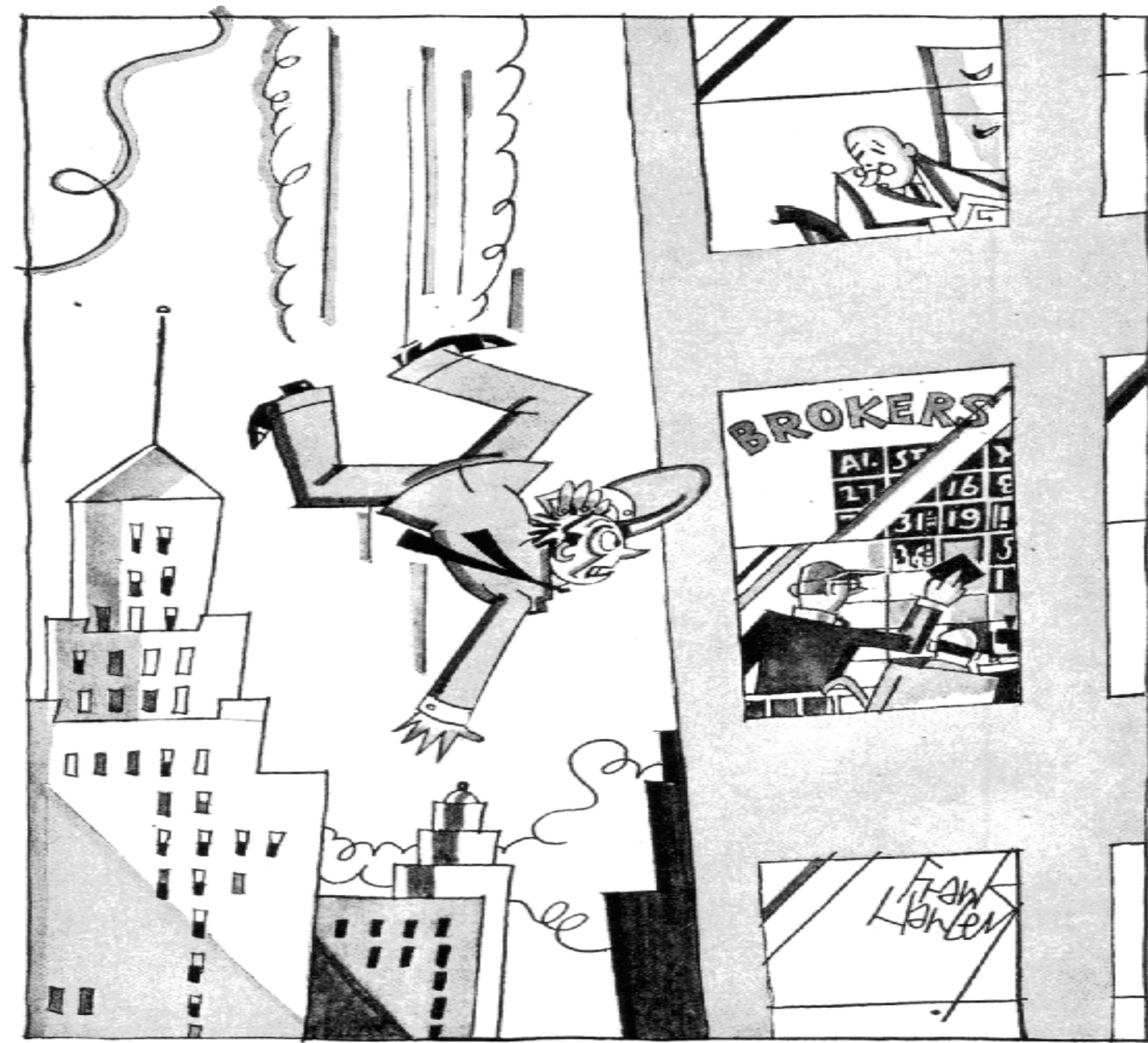
A massive amount of data gets generated daily hence predicting stock movement in a limited time is very challenging. We can use the Data Science processes to create machine learning models that can predict the stock price movement better than standard practices used in traditional strategies.

In this project, we research techniques that we can use to make investment decisions better. Good prediction models will offer a better understanding of stock price movement and associated risks. A good decision can impact their returns on equity.

Issues with Traditional Methods

Investing in stocks provides satisfactory returns if money is put in and withdrawn in an intelligent manner, but it involves a lot of risks. The Stock market is a very appealing and interesting area where it can reward you with a good return if strategies implemented works the way to wanted, at the same time if it doesn't work as expected it can cost you a lot of fortune.

There have been numerous attempts made to invent methods that can accurately predict stock price movement but no methods succeeded in providing 100% accurate results. There are many factors (i.e. Company performance, Industry Performance, Investor sentiments, and Economy – Interest rates, Economic outlook, Inflation, Deflation, Economic and political shocks, economic policy change, the value of currency) that affect stocks price and complexity of these attributes make it difficult to predict the market correctly. These attributes generate lots of data daily. Handling this huge amount of data increases the complexity of the task which is not easy to do through traditional methods. Also, stock markets are very dynamic and change in stock price happens within seconds. Traditional methods are not efficient at making decisions because they require a lot of time to process the information before they can make a decision.



Good preparation and market knowledge required for investment otherwise the same can happen to any investor

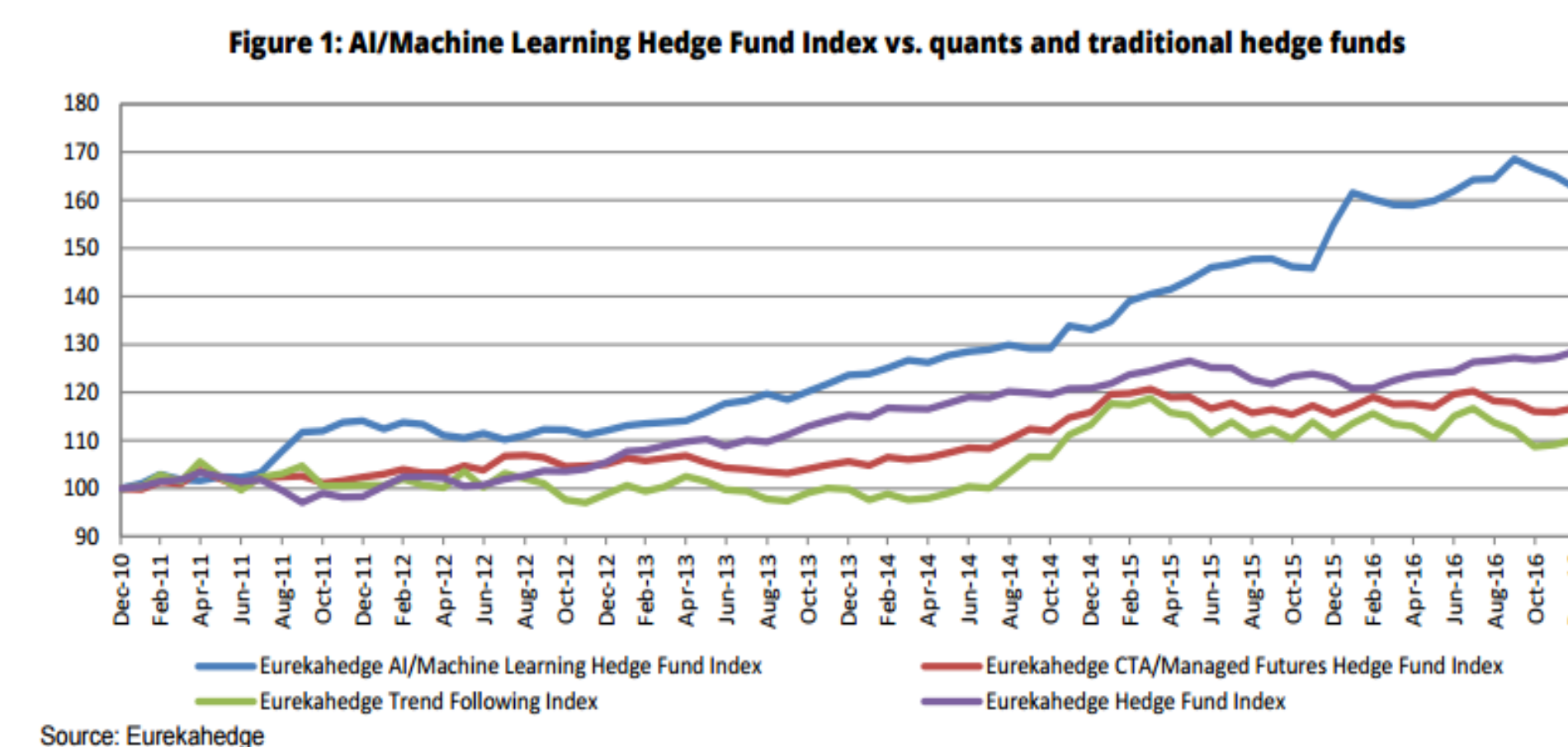
How Data Science can Help

Making profits in the stock market requires a lot of market knowledge and we need to learn how market behavior is based on past information to make the right judgment. Hence if we can build a 'historic-data-driven' predictive model that can predict probable market trends that such a model can be tremendously helpful for potential investors and would save them from the annoyance of analyzing data manually. This can be done using data science techniques. Data science techniques are very efficient at analyzing past data and retrieving insight from the data dumps.

Machine learning techniques can be used to analyze historical data related to stocks, economy, policy change, and industry to help understand how stocks have been performing in the past. This data can be used to predict how it's going to perform in the future. Also, this data can be tied with other information like real-time news to make the prediction even more accurate.

Online social networks like Twitter and Facebook are permitting people who are zealous about trading and investing to break critical financial news faster and they also go deeper into relevant areas of research and sources leading to real-time insights. We can use NLP (natural language processing) and different machine learning techniques to analyze sentiments and market news to predict the movement of stocks.

We can take the example of Eureka hedge funds. The firm Eureka hedge has published some interesting data. The graph below displays the performance of the Eureka hedge AI/Machine Learning Hedge Fund Index vs. traditional quant and hedge funds from 2010 to 2016 and we can see AI beats traditional funds.



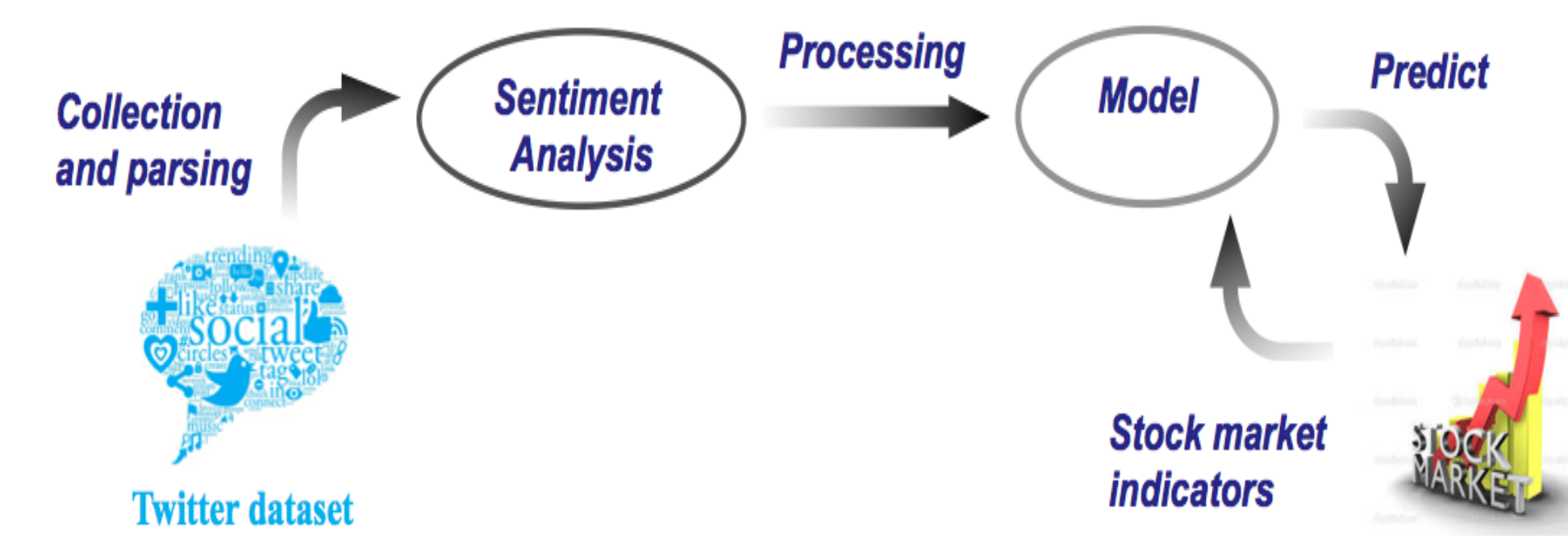
Eureka hedge states that "AI/Machine Learning hedge funds have outperformed both traditional quants and the average hedge fund since 2010, delivering annualized returns of 8.44% over this period compared with 2.62%, 1.62% and 4.27% for CTA's, trend-followers and the average global hedge fund respectively."

They also provided the details below on how AI hedge fund worked.

	Eureka hedge AI/Machine Learning Hedge Fund Index	Eureka hedge CTA/Managed Futures Hedge Fund Index	Eureka hedge Trend Following Index	Eureka hedge Hedge Fund Index
2011	14.10%	2.33%	0.71%	(1.75%)
2012	(1.80%)	2.66%	(1.66%)	7.34%
2013	10.34%	0.55%	1.02%	9.24%
2014	7.64%	9.66%	13.44%	4.89%
2015	16.40%	(0.31%)	(2.18%)	1.78%
2016	5.01%	1.15%	(0.62%)	4.48%
5 year annualised returns	7.35%	2.68%	1.80%	5.51%
5 year annualised volatility	4.95%	4.18%	7.13%	3.20%
5 year Sharpe Ratio (RFR=1%)	1.28	0.40	0.11	1.41
3 year annualised returns	9.57%	3.41%	3.31%	3.71%
3 year annualised volatility	5.61%	4.63%	7.78%	3.03%
3 year Sharpe Ratio (RFR=1%)	1.53	0.52	0.30	0.89
2 year annualised returns	10.56%	0.42%	(1.40%)	3.12%
2 year annualised volatility	6.31%	4.90%	8.07%	3.31%
2 year Sharpe Ratio (RFR=1%)	1.51	(0.12)	(0.30)	0.64

Data Science Techniques for Prediction

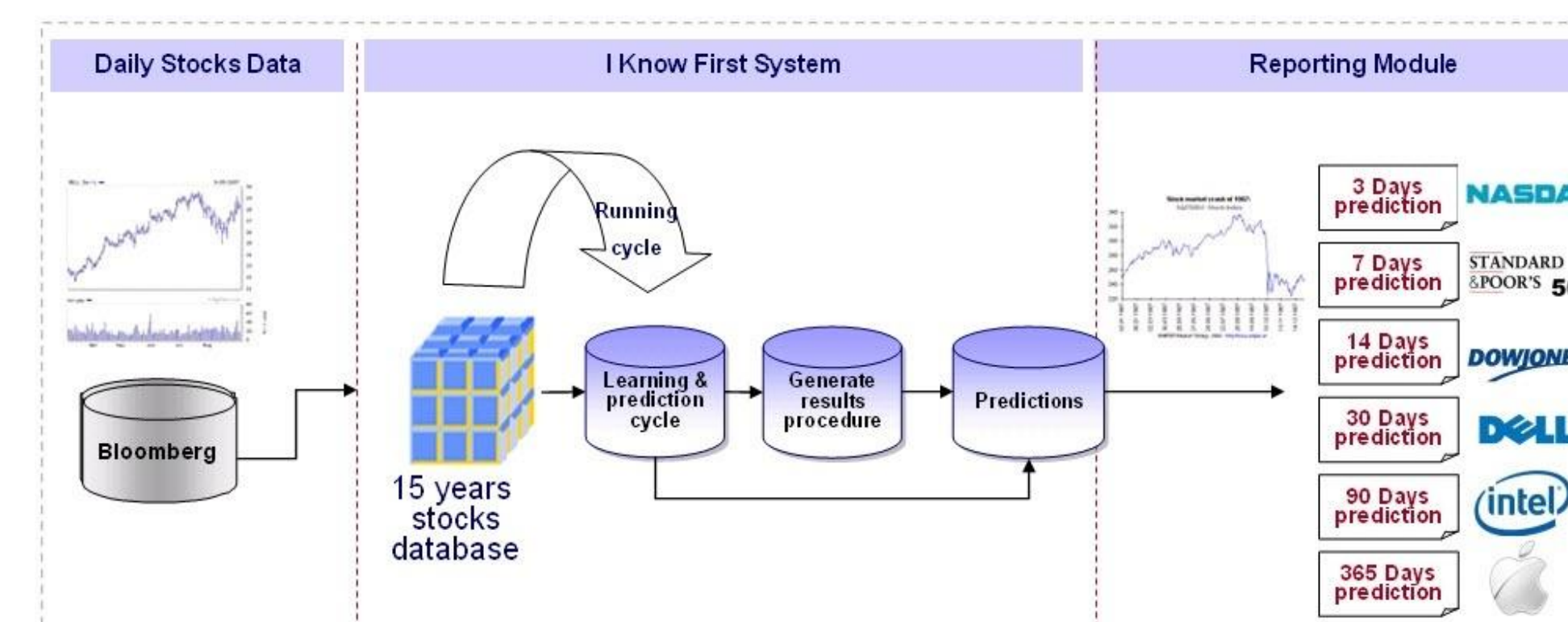
We can use Machine learning models and ingest sentiment data from trending news and social media to predict stock price movement. There are various techniques that can be utilized to understand market sentiments and predict accordingly. Natural language processing can be utilized to understand language data. Classification & regression methods or artificial neural networks can be used for prediction.



The picture shows an example of how we can use Social network data to predict stock price movement using data from Twitter.

Neural Networks: The most popular neural network training algorithm for financial forecasting is the back propagation (BP) algorithm, which is also a widely applied classical learning algorithm for neural networks. The BP neural network has been widely used in the area of financial time series forecasting because of its broad applicability to many business problems and its preeminent learning ability. Hybrid models of ANN are used as well to predict stock price.

Information like technical details, fundamental details, global economy details and Sentimental can be utilized all together to make predictions better. The below diagram shows how we can use historical data and learning techniques in predicting the market.



Results

The project provides Machine Learning models that can predict stock market movements using breaking news related to stocks from various social media networks like Twitter and Facebook.

Model efficiency and accuracy depend on training data ingested to train the model. If better data is supplied then the better model accuracy will be.



Conclusion

Based on the study done through this project I come to the conclusion below:

- It is difficult to predict stock price with 100% accuracy because of the complexity of the nature of attributes affecting stocks and the dynamic nature of the stock market.
- Data science techniques can help in predicting the direction of stock prices based on historical data and the latest market news.
- There is more work required to utilize daily market data, global economic data, and various other data related to stocks to make the stock prediction more accurate.
- A lot of investment is being done by investment companies like Goldman Sachs towards stock market prediction and related fields for innovation using Artificial Intelligence.
- A lot of new companies are coming up with a concept of Robo-Advisor for investment in Funds. The application of these kinds of advisors is increasing day by day.

References

- Hájek, P. (2018). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Computing & Applications*, 29(7), 343–358. Retrieved from <https://doi-org.ezproxy.bellevue.edu/10.1007/s00521-017-3194-2>
- Leekha, A., Wadhwa, A., Jain, N., & Wadhwa, M. (2018). Understanding the Impact of News on Stock Market Trends Using Natural Language Processing and Machine. *International Journal of Knowledge Based Computer Systems*, 6(2), 23–30. Retrieved from <http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=134635931&site=eds-live>
- Qiu, M., & Song, Y. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLoS ONE*, 11(5), 1–11. Retrieved from <https://doi-org.ezproxy.bellevue.edu/10.1371/journal.pone.015513>
- Machine Learning for Trading – Topic Overview <https://sigmoidal.io/machine-learning-for-trading/>
- Factors that can affect stock prices <https://www.getsmarteraboutmoney.ca/invest/investment-products/stocks/factors-that-can-affect-stock-prices/>
- DAS, D., SADIQ, A. S., AHMAD, N. B., & LLORET, J. (2017). Stock Market Prediction with Big Data Through Hybridization of Data Mining and Optimized Neural Network Techniques. *Journal of Multiple-Valued Logic & Soft Computing*, 29(1–2), 157–181. Retrieved from <http://ezproxy.bellevue.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=123624695&site=eds-live>

Acknowledgement

I thank professor Mr. Kenneth Birdsong for providing continuous constructive feedback and peers for their valuable inputs and suggestions. I also thank all the authors of the reference papers and articles.